PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Robert D. Palmquist	Confirmation No.	8120
Serial No.:	10/026,293	Customer No.:	28863
Filed:	December 21, 2001	Examiner:	Patrick Nestor Edouard
Docket No.:	1011-001US01	Group Art Unit:	2654
Title:	NETWORK-BASED TRANSLATION SYSTEM		

CERTIFICATE UNDER 37 CFR 1.8: I hereby certify that this correspondence is being deposited with the United States Post Service, as First Class Mail, in an envelope addressed to: Commissioner for Patents, Alexandria, VA 22313-1450 on May 16, 2006

By: Karen Sorensen
Name: Karen Sorensen

DECLARATION UNDER 37 C.F.R. 1.131

Commissioner for Patents
Washington, D.C. 20231

I, Robert D. Palmquist, declare as follows:

1. I am named sole inventor in above-referenced patent application serial no. 10/026,293.

2. I am an employee of Speechgear, Inc, the assignee of the above-referenced patent application.

3. As evidenced by this Declaration and the Exhibits referenced by this Declaration, I conceived the inventions set forth in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37 of this application prior to September 30, 2001, and worked diligently to reduce the inventions to practice from a time prior to September 30, 2001 through the time of actual reduction on or before November 26, 2001.

4. Furthermore, as evidenced by this Declaration and the Exhibits referenced by this Declaration, I conceived the inventions set forth in claims 6, 7, 9, 22 and 38 of this application prior to September 30, 2001, and worked diligently to reduce the inventions to practice from a time prior to September 30, 2001 through the time of constructive reduction to practice on December 21, 2001, i.e., the filing date of the current patent application.

Conception

5. Exhibit A, attached to this Declaration, includes excerpts of a progress report prepared for the United States government under SBIR Phase I Topic N01-044, Contract Number N00014-01-M-0225. The progress report from which the excerpts of Exhibit A were taken covered a two-week period of time that was prior to September 30, 2001, and was submitted to the United States government. Exhibit A specifically includes pages 2, 5-7 and 10-12 of the progress report.

6. Exhibit A provides evidence of my conception of the inventions set forth in the claims 1, 3, 4, 6, 7, 9, 16, 18, 22, 26, 28, 29, 32, 37 and 38 on or before the period covered by the progress report, which was before September 30, 2001. The actual dates in this Exhibit have been redacted.

7. Specifically, the pages marked as 10-12 of Exhibit A clearly demonstrate that the features recited in claims 1, 3, 4, 6, 7, 9, 16, 18, 22, 26, 28, 29, 32, 37 and 38 were in my possession at the time this progress report was prepared, which was prior to September 30, 2001.

8. Claim 1 recites a method comprising capturing an image containing text in a first language with a digital camera of a device, establishing with the device a wireless connection with a network, transmitting the image containing text in the first language from the device over the network via the wireless connection, receiving at the device a translation of the text in a second language over the network via the wireless connection, and displaying at the device the translation of the text in the second language. Conception of all of the features of claim 1 are supported in Exhibit A. Specifically, FIG. 5 of Exhibit A illustrates a device that includes a

digital camera. On page 10 of the status report, last line, to page 11 of the status report, first line, Exhibit A indicates that the digital camera will be used to capture an image of the foreign language. An example picture of this image captured by the digital camera is shown in FIG. 6. On page 11, lines 1-3, Exhibit A provides that a device establishes a wireless connection with a remote server and that the server translates the image and sends the translation back to the device. FIG. 7, on page 12 illustrates how the translated text is displayed on the device in the second language, e.g., along with a thumbnail representation of the original image. All of the features of claim 1 are clearly supported by Exhibit A, which was submitted to the government prior to September 30, 2001.

9. Claim 3 further requires display of the image. This is clearly supported in Exhibit A by FIGS. 6 and 7.

10. Claim 4 further requires displaying the image and displaying the translation of the text in the second language simultaneously. This is clearly supported in Exhibit A by FIG. 7.

11. Claim 6 further requires transmitting a second image containing second text in the first language over the network, and receiving a translation of the first text and the second text in the second language over the network. Claim 7 further requires transmitting the first image and the second image over a network in response to a single command from a user. Claims 6 and 7 are clearly supported by Exhibit A on page 11, lines 13-41, which discuss the "one-click" option for sending multiple images.

12. Claim 9 recites compressing the image. Exhibit A discloses this feature at page 11, lines 5-9.

13. Claim 12 recites that the network comprises a cellular telephone network Exhibit A discloses this feature at page 11, lines 4-5.

14. Claim 16 recites a device comprising a digital camera that captures an image containing text in a first language, a transmitter that transmits the image over a network via a wireless connection so that the text can be translated by a different device, a receiver that receives a translation of the text in a second language over the network via the wireless connection, and a display that displays the translation of the text in the second language. Conception of all of the features of claim 16 are supported in Exhibit A. Specifically, FIG. 5 of Exhibit A illustrates a device that includes a digital camera. On page 10, last line, to page 11, first line, Exhibit A indicates that the digital camera will be used to capture an image of the foreign language. An example picture of this image captured by the digital camera is shown in FIG. 6. On page 11, lines 1-3, Exhibit A provides that device establishes a wireless connection with a remote server and that the server translates the image and sends the translation back to the device. FIG. 7, on page 12 illustrates how the translated text is displayed on the device in the second language, e.g., along with a thumbnail representation of the original image. All of the features of claim 16 are clearly supported by Exhibit A, which was submitted to the government prior to September 30, 2001.

15. Claim 18 requires the simultaneous display of the translation and the image. This is clearly supported in Exhibit A by FIG. 7.

16. Claim 22 requires that the device comprise a cellular telephone that establishes the wireless connection so that the text can be translated by the different device. Exhibit A discloses this feature at page 11, lines 4-5.

17. Claim 26 is an independent claim to a system comprising a client device including a digital camera that captures an image containing text in a first language, a client transmitter that transmits the image over a network to a remote server via a wireless connection so that the text can be translated by the remote server, a client receiver that receives a translation of the text in a second language over the network from the remote server via the wireless connection, and a display that displays the translation of the text in the second language; and the remote server including a receiver that receives the image over the network from the client device, a translator

that generates the translation of the text in the second language and a transmitter that transmits the translation over the network to the client device. Conception of all of the features of claim 16 are supported in Exhibit A. As noted above with respect to claims 1 and 16, Exhibit A illustrates a device that includes a digital camera. On page 10, last line, to page 11, first line, Exhibit A indicates that the digital camera will be used to capture an image of the foreign language. An example picture of this image captured by the digital camera is shown in FIG. 6. On page 11, lines 1-3, Exhibit A provides that a device establishes a wireless connection with a remote server and that the server translates the image and sends the translation back to the device. FIG. 7, on page 12, illustrates how the translated text is displayed on the device in the second language, e.g., along with a thumbnail representation of the original image. All of the features of claim 26 are clearly supported by Exhibit A, which was submitted to the government prior to September 30, 2001.

18. Claim 28 is a method claim that requires capturing a first image containing text in a first language with a digital camera of a device, generating from the first image a second image containing the text in response to a command from a user, wherein generating the second image includes editing out one or more portions of the first image that do not include the text, transmitting the second image from the device over a network so that the text can be translated, receiving at the device a translation of the text in a second language over the network, and displaying at the device the second image and the translation. Claim 28 is also supported by Exhibit A at pages 10-12 and FIGS. 5-7. The passages addressed above with respect to claims 1, 16 and 26 also provide support for the features of claim 28.

19. Claim 29 further requires establishing the wireless connection with the network, which is clearly supported in Exhibit A at page 11, lines 2-3.

20. Claim 32 requires display of the second image and the translation simultaneously. As noted above, this is clearly supported in Exhibit A by FIG. 7.

21. Claim 37 recites a method comprising capturing an image containing text in a first language with a digital camera of a device, transmitting the image containing text in a first language from the device over a network so that the text can be translated, receiving at the device a translation of the text in a second language over the network, and displaying the image and the translation simultaneously at the device. Claim 28 is also supported by Exhibit A at pages 10-12 and FIGS. 5-7. The passages addressed above with respect to claims 1, 16 and 26 also provide support for the features of claim 28.

22. Claim 38 recites storing a plurality of images containing text, and transmitting at least a portion of the plurality of images over the network in response to a single command from a user. Claim 38 is clearly supported by Exhibit A on page 11, lines 13-41, which discuss the "one-click" option for sending multiple images.

Reduction to Practice

23. Exhibit B, attached to this Declaration, provides evidence that the inventions recited in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37 were actually reduced to practice on or before November 26, 2001.

24. The filing of the current application was a constructive reduction to practice of the features recited in claims 6, 7, 9, 22 and 38, i.e., on December 21, 2001.

25. Exhibit B contains excerpts from another progress report prepared for the United States government under SBIR Phase I Topic N01-044, Contract Number N00014-01-M-0225. The progress report from which the excerpts of Exhibit B were taken covered a period between November 10, 2001 and December 10, 2001. Exhibit B specifically includes pages 2, 5-7 and 14-18 of the progress report.

26. Page 7 of the progress report in Exhibit B provides a status overview. In this status overview on page 7, the progress report indicates that a successful demonstration of an

English-to-Arabic system was given to the Office of Naval Research on November 26, 2001. Page 7 of the progress report in Exhibit B specifically indicates that the demonstration included the "camera-based" mode, which is discussed in pages 14-18 of the progress report.

27. Pages 14-18 of the progress report in Exhibit B specifically illustrate the functionality of the prototype covered by claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37.

28. Specifically, the features of independent claims 1, 16, 26, 28 and 37 were clearly supported by the demonstrated prototype, as discussed on pages 14-18 of Exhibit B. FIGS. 11-14 on pages 17 and 18 specifically illustrate operation of the prototype. FIG. 11 illustrates an image that includes Arabic text, which was captured by a digital camera of the device. FIG. 12 illustrates a user selection to translate the text in the image. FIG. 13 illustrates the device uploading the image to a server. FIG. 13 also illustrates a toolbar to show progress of a subsequent translation download from the server. FIG. 14 illustrates the final translation of the Arabic text as "Post Office." Clearly, the features of claims 1, 16, 26, 28 and 37 were reduced to practice on or before November 26.

29. The features of claims 3 are shown as being reduced to practice in FIG. 11 of Exhibit B, which shows display of the image including Arabic text to be translated.

30. The features of claims 4, 18 and 32 are shown as being reduced to practice in FIG. 14 of Exhibit B, which shows display of the image including Arabic text to be translated simultaneously with the display of the translation of this text into English, i.e. "Post Office."

31. The features of claim 29, i.e., establishing the wireless connection, are shown as being reduced to practice in FIG. 13 of Exhibit B, which shows progress of the image download process.

Diligence

32. With respect to the inventions recited in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37, during the period from prior to September 30, 2001 to the actual reduction to practice on November 26, 2001, Speechgear worked diligently toward the reduction to practice.

33. Exhibit C is an excerpt from a progress report covering a period between September 10, 2001 and November 9, 2001. Exhibit C specifically includes pages 2, 6 and 13-15 of the progress report.

34. Exhibits C and B demonstrate the diligence over a period between a date prior to September 30, 2001 and the date the inventions recited in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37 were actually reduced to practice, i.e., on or before November 26, 2001.

35. The progress report associated with Exhibit C was submitted after the progress report associated with Exhibit A, which is currently being used to establish conception of the claimed inventions prior to September 30. During the period covered by Exhibit C, Speechgear worked diligently towards reducing the invention to practice, specifically for purposes of demonstrating a prototype to the government.

36. Over the period covered by the progress associated with Exhibit C from a time prior to September 30, 2001 to November 9, 2001, Speechgear employees worked substantially every day (not necessarily including weekends or holidays) on advancing the project towards a reduction to practice of the features recited in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37.

37. Also, over the period covered by the progress report associated with Exhibit B from November 10, 2001 to the reduction to practice of the invention on or before November 26, Speechgear employees continued working diligently substantially every day (not necessarily including weekends or holidays) on advancing the project towards a reduction to practice.

38. With respect to the inventions recited in claims 6, 7, 9, 22 and 38, during the period from prior to September 30, 2001 to the constructive reduction to practice on December 21, 2001 Speechgear worked diligently toward the reduction to practice.

39. Claims 6, 7, 9, 22 and 38 are dependent claims. With respect to these claims, the activities of Speachgear addressed above to demonstrate diligence with respect to the independent claims, and therefore also apply to claims 6, 7, 9, 22 and 38 insofar as these claims incorporate all the features of their respective independent claims.

40. Following the diligence to the actual reduction to practice of inventions recited in claims 1, 3, 4, 16, 18, 26, 28, 29, 32 and 37 on or before November 26, 2001, Speechgear continued working diligently in improving the prototype to include other features, including those recited in claims 6, 7, 9, 22 and 38.

41. Over the period covered by the progress associated with Exhibit B through December 10, Speechgear employees worked substantially every day (not necessarily including weekends or holidays) on advancing the project towards a reduction to practice, including work on the features recited in claims 6, 7, 9, 22 and 38.

42. Exhibit D is another progress report excerpt, associated with a report covering the period of December 11, 2001 to January 8, 2001. Exhibit D includes pages 2, 6 and 16-21 of the progress report.

43. Over the period covered by the progress report associated with Exhibit D through the constructive reduction to practice of the features recited in claims 6, 7, 9, 22 and 38 on December 23, Speechgear employees worked substantially every day (not necessarily including weekends or holidays) on advancing the project towards a reduction to practice.

44. On pages 18-19, the progress report associated with Exhibit D specifically identifies progress over this period in advancing details of system requirements related to the features recited in claims 6, 7, 9, 22 and 38.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true. I further declare that these statements are being made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Date: 15 May 06 Signed: 

Robert D. Palmquist

Progress Report

Project
- ***Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution***
- SBIR Phase I Topic N01-044
- Contract Number N00014-01-M-0225

Item Number
0001AA: Progress Report

Security Classification
Unclassified

Sponsor
Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Principle Investigator:
Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

SPEECH GEAR

Program Partners

VISUAL GOLD

SDL INTERNATIONAL

ARAMEDIA

Compadre:
A Device Independent Voice-to-Voice Language Translator Software Solution

---

A Phase I SBIR Speech Application Project
for the Office of Naval Research

SBIR Call Number N01-044
Contract Number N00014-01-M-0225

SPEECH GEAR

Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

Telephone: 507-664-0974
Email: robert@palmquist.com
Web: www.speechgear.com

A. Project Summary

Technical Abstract:

Mission Statement
To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

B. Project Status

B.1 Status Overview:

The overall work breakdown structure is provided in Figure 1. For purposes of this report, the project start date is selected at . The actual purchase order has not yet been received in the mail, however a FAX copy of the signed document was provided by Jennifer Schoen on

As is shown in Figure 1, the project is currently ahead of schedule. Prospective Users of the system have been interviewed and the resulting Design Requirements (DR) has been drafted. This document is included in this report as Appendix A. The DR contains the targeted and desired specifications for *Compadre's* overall system performance.

The system is divided into three basic areas: standalone, camera-based and telephone-based operations. These three areas are also listed in order of difficulty, with the standalone mode being the easiest to implement and the telephone based system being most difficult. Progress has been made in each of these three categories. This progress is described in the remainder of Section B.

B.2 StandAlone Mode

In this mode, the PDA/Cellphone (henceforth called a "SmartPhone") will not be required to wirelessly connect to a remote server. The translation capabilities will be primarily bi-directional word look-up. Initially, the interface will be a touchscreen such as is shown in Figure 2. Multiple language pairs will be supported along with a 30,000+ word dictionary.

Six different vendors have been identified for potential teaming partners on developing *Compadre*: AIM, Smart Link, TomTom, Evolutionary Systems, PhatWare and Ectaco. Each of these vendors have provided samples of their current software product, and these are in the process of being evaluated. The templates for this two-part evaluation are provided in Figures 3 and 4. The first template

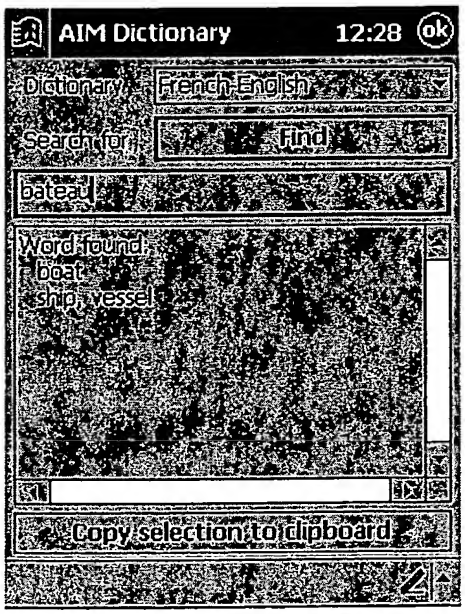

Figure 2: Example of Touchscreen Interface for Stand-Alone Mode

	HPC Translate	Pocket Context	Pocket Language Teacher	Collins Dictionaries	Dictionary	Travel Dictionary
Vendor	PhatWare	SmartLink	Ectaco	TomTom	Evolutionary Systems	AIM
Size of English Dictionary						
Ability to Add User Specific Terminoligy						
Bi-Directional Capability						
Ease of Use						
Number of Languages Supported						
List of Languages						
List of Supported CPUs						
Additional Comments						

Figure 4: Part 2 of 2 for Evaluating Stand-Alone Translator Products

will be used to test specific word translation capabilities. The second template is used to evaluate overall system capabilities. Based on the results of these tests, two vendors will be selected as partners to continue development activities.

B.3 Camera-Based Mode

The primary means to input text into the SmartPhone for this mode of usage will be a digital camera. A patent application for this capability has been submitted. Such a system is shown in Figure 5. The digital camera will be used to capture an image of the

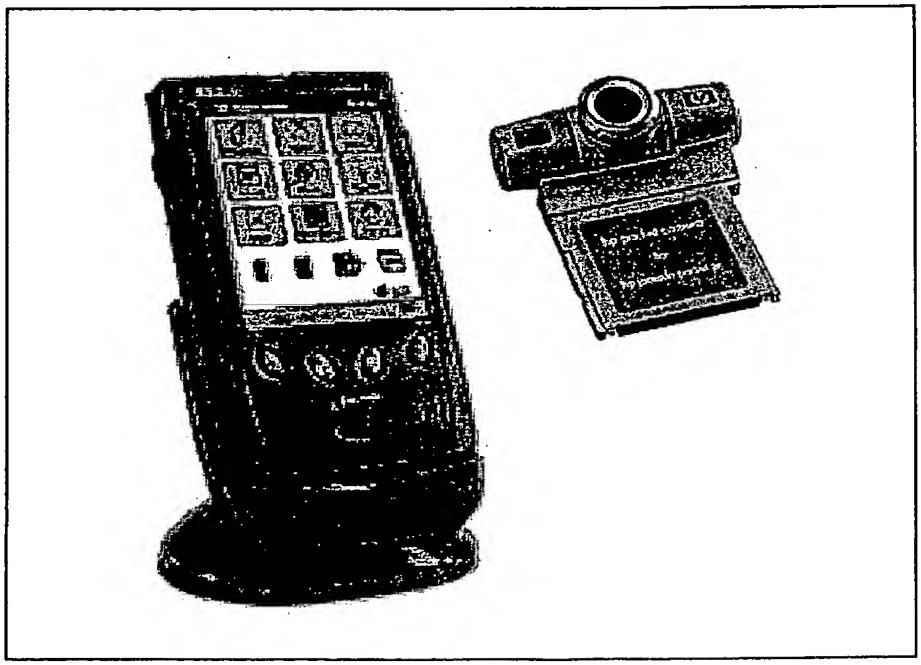

Figure 5: Example of Camera-based System

foreign language. Such a picture is shown in Figure 6[1]. Once the desired image is obtained, the SmartPhone will wirelessly connect to a remote server where the image will be processed and the resulting translation sent back to the user. An example of the translated text in the proposed "one-click" GUI is shown in Figure 7. For most applications, this connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear is in the process of evaluating different image compression algorithms. These algorithms will be embedded directly into SpeechGear's software, and thus will be transparent to the end user. The current plans are to use Visual Gold's *Imagist* product. This can be viewed at www.visualgold.com. SpeechGear has had initial meetings with Visual Gold and the appropriate NDA's have been signed. A "Letter of Intent" with respect to the teaming arrangement is in the process of being drafted.

As is shown in Figure 6, a "one-click" GUI is planned. After capturing the image(s), the user will simply select "Translate" and the wireless connection will automatically be established. Note that multiple images can be sent simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, items that are selected can be loaded into a virtual basket or cart, and once you are done shopping you can select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll

Figure 6: Preliminary Functional Layout of Graphical User Interface – Text Boxes will be Replaced with Icons

[1] Note, since our software is not yet functional, I have no idea what this Arabic text says. If the option is exercised, we will be hiring an Arabic speaking individual to be part of our team.

through these images and save or delete them as is desired. Please note that the actual buttons will be Icons versus text, and thus the look and feel of the resulting GUI will be a substantial improvement over what is shown in the Figures.

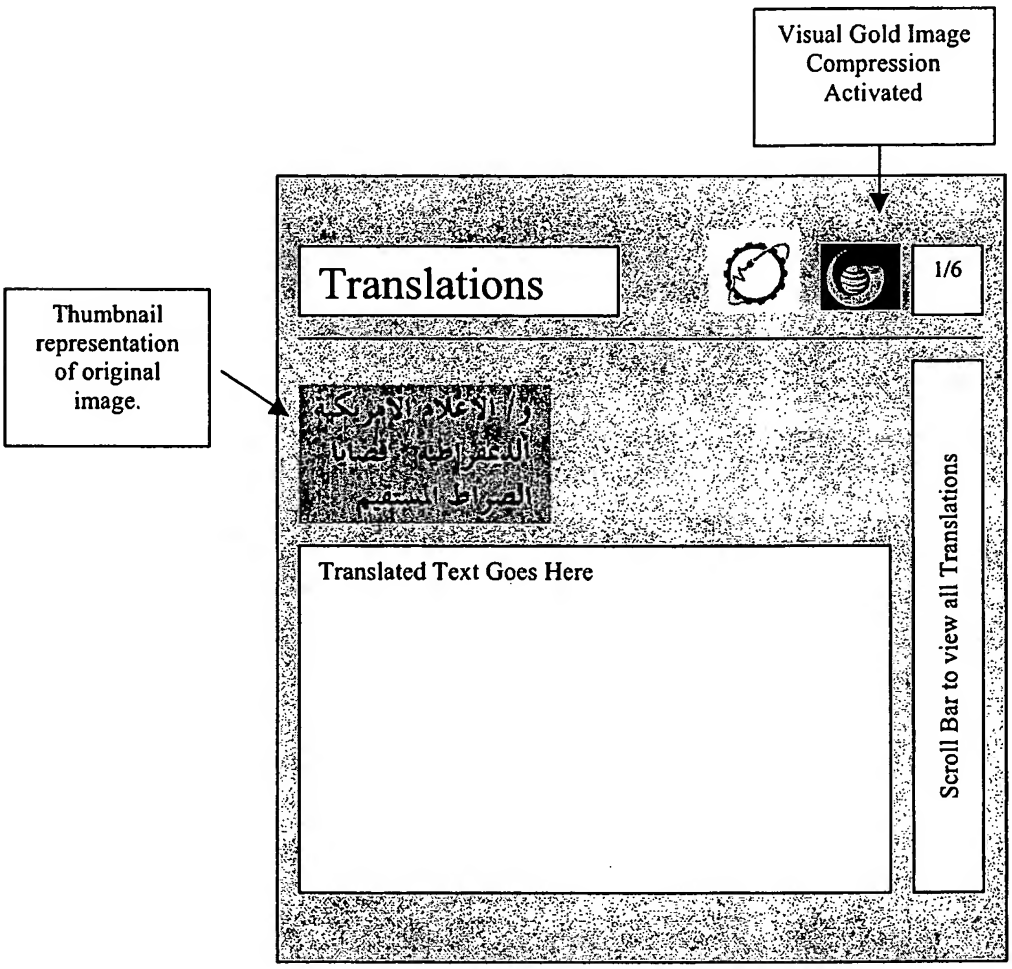

Figure 7: Preliminary Graphical User Interface

Progress Report

Period Covered by the Report
• *10 November 2001 to 10 December 2001*

Date of Report
• *10 December 2001*

Project
• ***Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution***
• SBIR Phase I Topic N01-044
• Contract Number N00014-01-M-0225

Item Number
0001AD: Progress Report

Security Classification
Unclassified

Sponsor
Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Principle Investigator:
Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

SPEECH GEAR

Program Partners

VISUAL GOLD | IRISUSA | ARAMEDIA | L&H/Apptek | CiMOS

Compadre:
A Device Independent Voice-to-Voice Language Translator Software Solution

---

A Phase I SBIR Speech Application Project
for the Office of Naval Research

SBIR Call Number N01-044
Contract Number N00014-01-M-0225

SPEECH GEAR

Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

Telephone: 507-664-9123
Email: rpalmquist@speechgear.com
Web: www.speechgear.com

A. Project Summary

Technical Abstract:

Mission Statement
To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

B. Project Status

B.1 Status Overview:

The overall work breakdown structure is provided in Figure 1. For purposes of this report, the project start date is selected at The actual purchase order was not received in the mail, however a FAX copy of the signed document was provided by Jennifer Schoen on

As is shown in Figure 1, a successful demonstration of the English/Arabic proof-of-concept system was given at the Office of Naval Research on November 26, 2001. This included all three usage modes: standalone, camera-based and voice-based. The demonstrations were performed commensurate with the Design Requirements (DR) and Prototype System Design (PSD) documents that were developed during the course of this Phase I effort with the only exception being that the voice-based system was demonstrated using a laptop versus using telephones to connect to a remote server. The DR, which is included in Appendix A of this report, contains the targeted and desired specifications for *Compadre's* overall system performance. This document was submitted in the July progress report and was approved per telephone conversations with Dr. Joel Davis. The PSD document, which is included in Appendix B of this report, contains a description of the overall system design. This document was submitted in the September progress report and was subsequently approved. In short, the DR describes what the system does, whereas the PSD describes how this is accomplished. The one critical item that remains is to use a telephone to collect spoken phrases versus a microphone headset. The required hardware (e.g., TAPI modem) has been evaluated, procured and installed. The software components have also been either acquired or written. Work is continuing to achieve this capability with a targeted completion date of December 24, 2001.

B.2.2 Camera-Based Mode

There are situations where using a touchscreen or keyboard to input foreign text will not be practical. One such example is the sign containing Arabic text that is shown in Figure 8. In this situation, it would be very difficult for an English-only speaking individual to enter the Arabic text using a keyboard or touchscreen or to look-up this text in a traditional English/Arabic dictionary. The same situation is present for multiple languages such as Korean, Japanese and Russian. To help solve this problem, *Compadre* allows the user to input text into the SmartPhone using a digital camera. A patent application for this capability has been submitted. The design of the prototype system is shown in Figure 9. Two different cameras are being used: a compact camera from HP that is very convenient to use and a high resolution camera from Minolta with superior capabilities but a more involved interface. The Minolta *Dimage 7* is being used to develop translation capabilities for full text documents with small font sizes (e.g., a complete page of Arabic text) whereas the HP camera is used for larger font sizes such as signs.

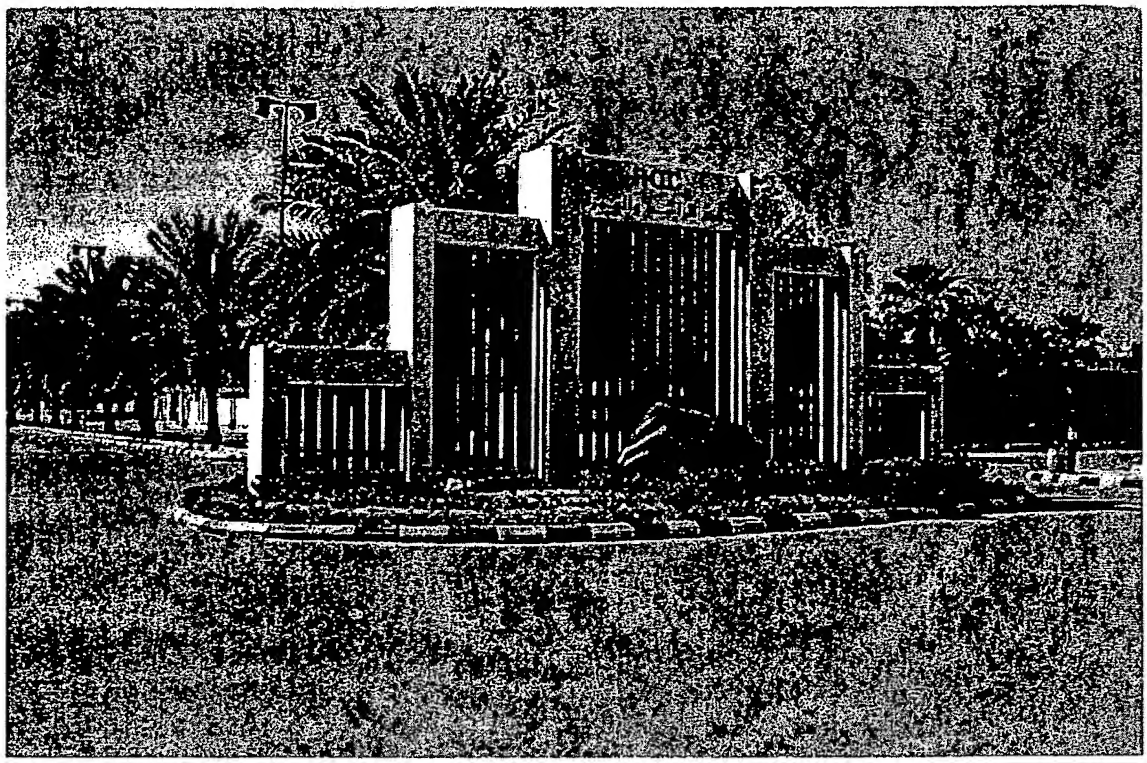

Figure 8: Examples of Arabic Sign

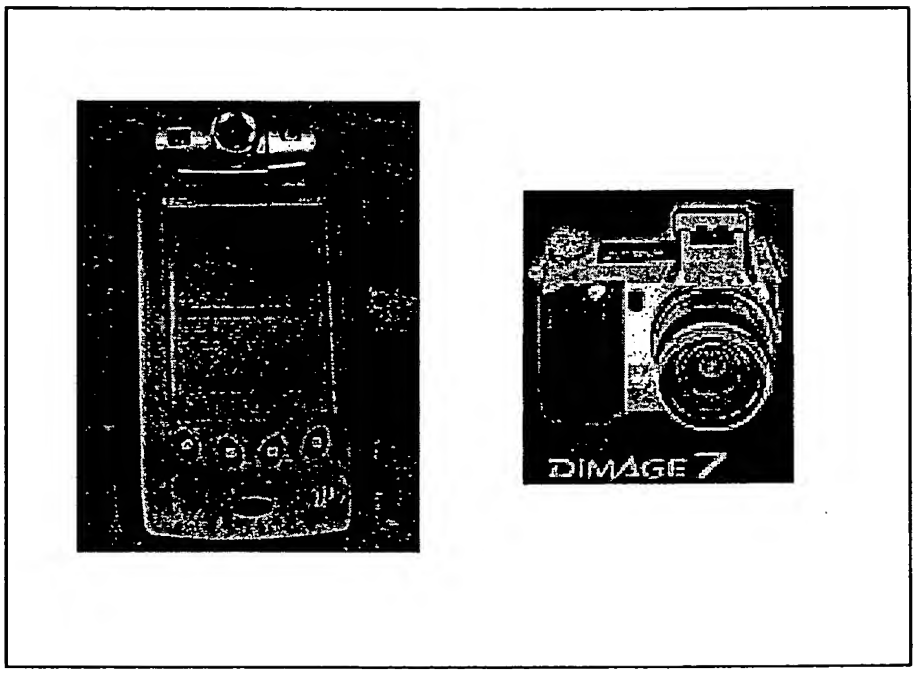

Figure 9: Examples of Camera-based Systems

Note that *Compadre's* software is designed to be device independent, thus, these are just two of many hardware configurations that could be used for this usage mode. One interesting alternative device is Samsung's conceptual product of including a camera with a cellular phone. This product is shown in Figure 10.

The digital camera is used to capture an image of the foreign language as is shown in Figure 11. Once the desired image is obtained, a "one-click" GUI is used to wirelessly connect the SmartPhone to a remote server where the image will be processed and the resulting translation sent back to the user. This is shown in Figure 12. This process takes approximately one minute to complete with the vast majority of this time being consumed by uploading the image to the server. Status bars, which are shown in Figure 13, are displayed to inform the user as to the percentage completion of each of the uploading and downloading procedures. The resulting translation is then provided along with the original picture. An example of this is shown in Figure 14. Note that for most situations the wireless connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear evaluated different image compression algorithms and selected the *Imagist* product from Visual Gold. SpeechGear is currently embedding *Imagist* directly into SpeechGear's software. This, along with several other features SpeechGear will implement in Phase II, will significantly reduce the time it takes to

upload images and thus reduce the overall time it takes to complete the translation process.

An additional user screen is accessed by selection the "tools" tab, which is located at the bottom of the user interface (see Figure 14). This screen, which is shown in Figure 15, is used to specify parameters, such as the host address, user account and password, of the remote server that *Compadre* is using to perform the translation process. Individuals can use this tool in the field to establish connectivity with additional severs. For example, if the a laptop is residing in a vehicle, or a soldier's has a wearable computer, the user could redirect the connectivity to this nearby platform and use Infrared or 802.11 to provide the connectivity versus a cellular telephone.

For the Phase I proof-of-concept system, the following phrases, in Arabic, have been included in the system:
"Hospital"
"Speed Limit 50"
"No Parking"
"Grocery Store"
"Post Office"
"Telephone"
"Emergency Use Only"
"Authorized Personnel Only"
"Danger, Do Not Enter"
This set of possible signs was selected to place a boundary on the overall scope of the OCR software requirements. In Phase II this limitation of preselected phrases will be removed.

Figure 10: Samsung's Proposed Combined Camera and Digital Cellular Phone

Currently only one image can be sent at a time. However, in the future the user will be able to send multiple images simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, selected items are loaded into a virtual basket or cart, and once you are done shopping you select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll through these images and save or delete them as is desired.

One item of note is that *Compadre's* Hybrid Translator can be configured to handle different types of input using a variety of methods. For voice-based input, the context in which words are used is readily available. This often is not the case with the camera-

based mode. For example, the words "Post Office" without context could be interpreted as a "Pole that is stuck in the ground" and "A place where people work." Thus, SpeechGear configured the translator to be dominated by a Translation Memory (TM) mode versus Machine Translation (MT). In TM, the translator uses a known set of previously translated phrases to achieve accurate outputs. Such an approach is used very often if for example an operator's manual has been previously translated, but has now been updated and thus needs to be translated once again. In the case of the camera-based system, the TM approach will be used to enter signs and information, such as the Post Office example that was stated above. Thus, SpeechGear is in the process of building the TM database to include signage typically seen on signs.

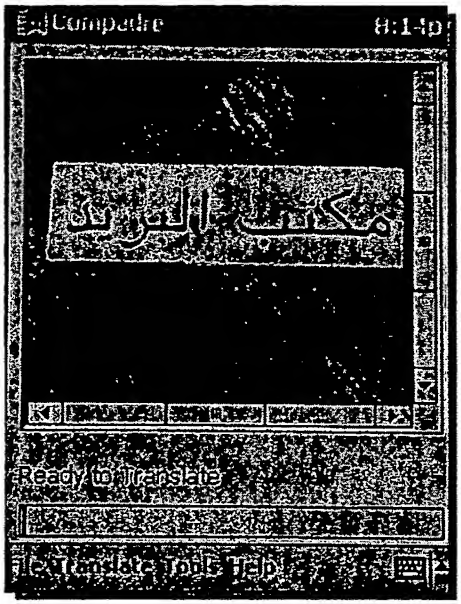

Figure 11: Example of Touchscreen Interface for Stand-Alone Mode

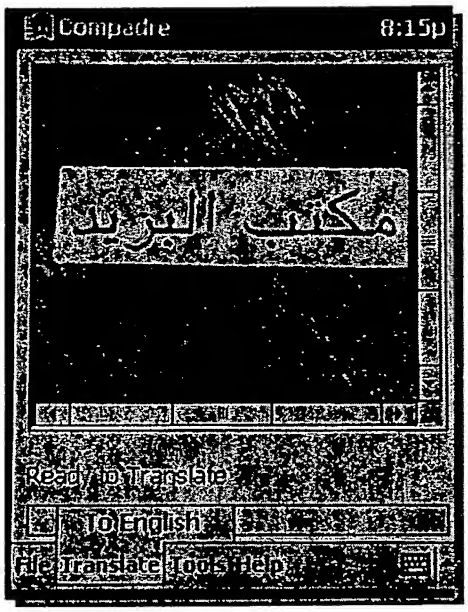

Figure 12: Example of Touchscreen Interface for Stand-Alone Mode

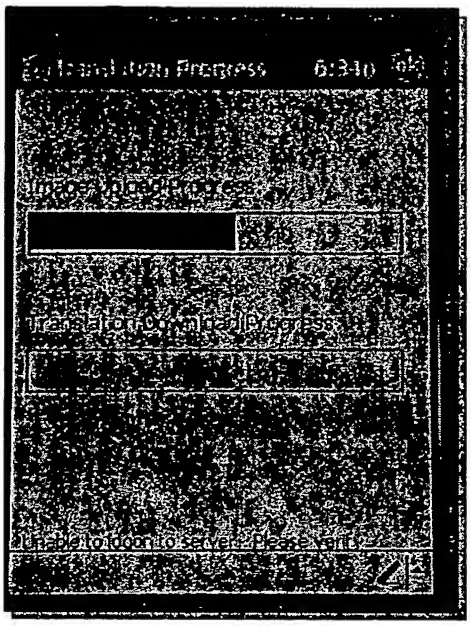

Figure13: Graphical User Interface for Viewing Results of Translation

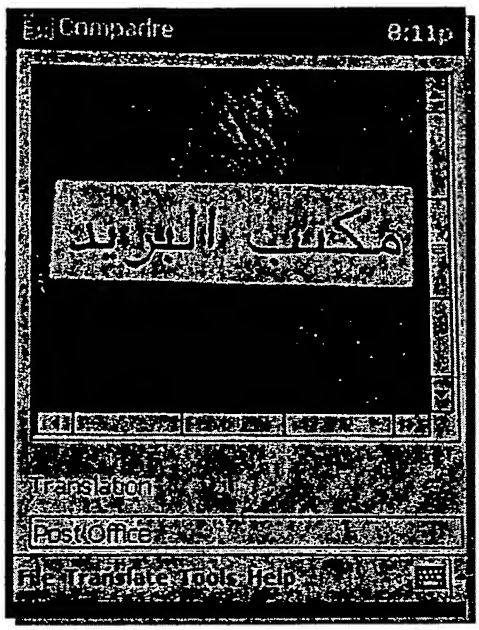

Figure14: Graphical User Interface for Viewing Results of Translation

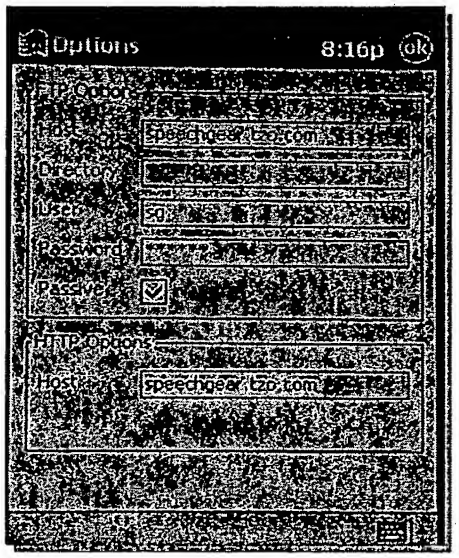

Figure15: Graphical User Interface for Viewing Results of Translation

Progress Report

Period Covered by the Report
• *10 September 2001 to 9 November 2001*

Date of Report
• *10 November 2001*

Project
• ***Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution***
• SBIR Phase I Topic N01-044
• Contract Number N00014-01-M-0225

Item Number
0001AC: Progress Report

Security Classification
Unclassified

Sponsor
Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Principle Investigator:
Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

SPEECH GEAR

Program Partners

VISUAL GOLD | IRISUSA | ARAMEDIA | L&H/Apptek | CiMOS

A. Project Summary

Technical Abstract:

Mission Statement
To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

B.3 Camera-Based Mode

The primary means to input text into the SmartPhone for this mode of usage will be a digital camera. A patent application for this capability has been submitted. The design of the prototype system is shown in Figure 5. Two different cameras are being used: a compact camera from HP and a high resolution camera from Minolta. The Minolta *Dimage 7* is being used to perform the initial testing for *Compadre*. Once this camera has been successfully integrated and tested, then SpeechGear will proceed to integrate and test lower resolution cameras such as the HP camera that is shown.

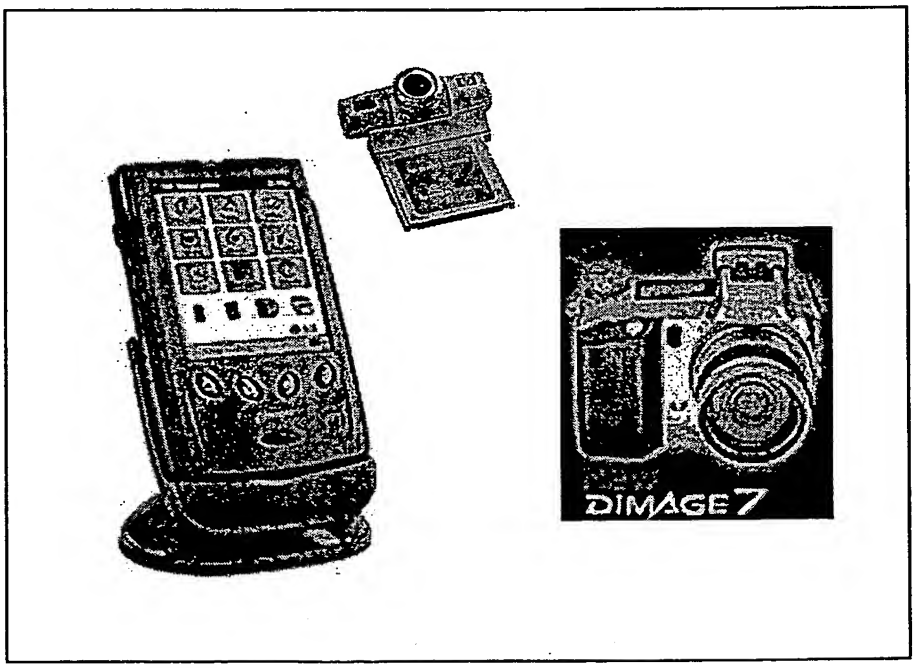

Figure 5: Examples of Camera-based Systems

Note that *Compadre's* software is designed to be device independent, thus, these are just two of many hardware configurations that could be used for this usage mode. One interesting alternative device is Samsung's conceptual product of including a camera with a cellular phone. This product is shown in Figure 6.

The digital camera will be used to capture an image of the foreign language. Such a picture is shown in Figure 7. Once the desired image is obtained, the SmartPhone will wirelessly connect to a remote server where the image will be processed and the resulting translation sent back to the user. An example of the translated text in the "one-click" GUI is shown in Figure 7. For most applications, this connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear evaluated

different image compression algorithms and selected the *Imagist* product from Visual Gold. *Imagist* will be embedded directly into SpeechGear's software, and thus will be transparent to the end user.

The GUI being developed for *Compadre* is shown in Figure 7. After capturing the image(s), the user will simply select "Translate" and the wireless connection will automatically be established. Note that multiple images can be sent simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, selected items are loaded into a virtual basket or cart, and once you are done shopping you select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll through these images and save or delete them as is desired. Please note that the actual buttons will be Icons versus text, and thus the look and feel of the resulting GUI will be a substantial improvement over what is shown in the figures.

Figure 6: Samsung's Proposed Combined Camera and Digital Cellular Phone

One item of note is that *Compadre's* Hybrid Translator can be configured to handle different types of input using a variety of methods. For voice-based input, the context in which words are used is readily available. This often is not the case with the camera-based mode. For example, the words "Post Office" without context could be interpreted as a "Pole that is stuck in the ground" and "A place where people work." Thus, SpeechGear is configuring the translator to be dominated by a Translation Memory (TM) mode versus Machine Translation (MT). In TM, the translator uses a known set of previously translated phrases to achieve accurate outputs. Such an approach is used very often if for example an operator's manual has been previously translated, but has now been updated and thus needs to be translated once again. In the case of the camera-based system, the TM approach will be used to enter signs and information, such as the Post Office example that was stated above. Thus, SpeechGear is in the process of building the TM database to include signage typically seen on signs.

Figure 7: Preliminary Graphical User Interface to Submit Images for Translation

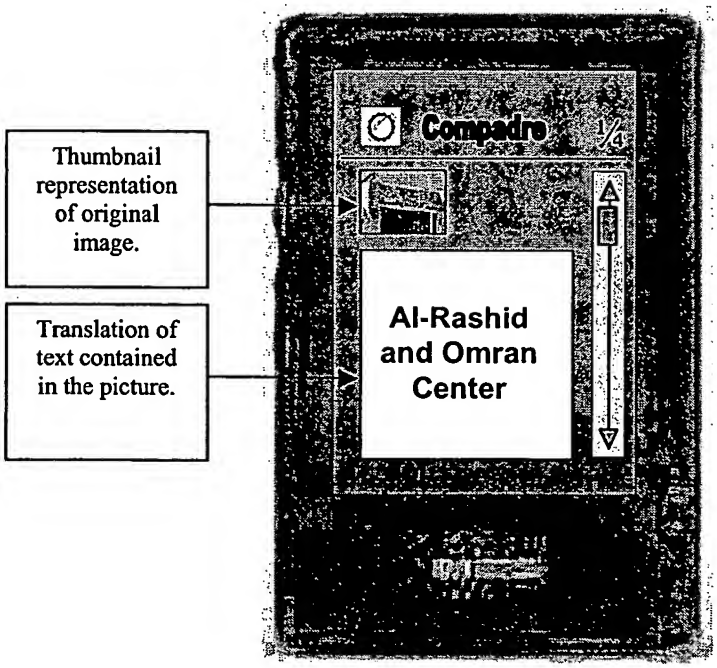

Figure 8: Preliminary Graphical User Interface for Viewing Results of Translation

Progress Report

Period Covered by the Report
• *11 December 2001 to 8 January 2002*

Date of Report
• *9 January 2002*

Project
• ***Compadre: A Device Independent Voice-to-Voice Language Translator Software Solution***
• SBIR Phase I, Topic N01-044
• Contract Number N00014-01-M-0225, Amendment P00001

Item Number
0002AA: Progress Report

Security Classification
Unclassified

Sponsor
Dr. Joel Davis
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, VA 22217-5660

Principle Investigator:
Robert Palmquist
SpeechGear, Inc
3454 Circle Bluff Court
Faribault, MN 55021

SPEECH GEAR

Program Partners

VISUAL GOLD | IRISUSA | ARAMEDIA | L&H/Apptek | MSU

A. Project Summary

Technical Abstract:

Mission Statement
To develop and deploy language translation software that is device independent, supports bi-directional translation of multiple languages, produces text transcriptions of spoken conversations and supports translation of text extracted from digital images. This software shall run in both a reduced functionality standalone mode, and by wirelessly connecting to remote servers, a full-function mode. This software shall run on multiple pocketable platforms resulting in a mobile system that is low in cost, easy to use, robust in operation and comfortable to carry and/or wear.

The object of this Phase I research effort is to investigate the scientific, technical and commercial merit and feasibility of the system described in the preceding mission statement. Specifically, the team will investigate design options for the mobile translator system, identify potential applications, and select the best option(s) to pursue in making the design a reality. Four technical areas will be investigated: potential pocketable computing platforms, the operator interface, optical character recognition software and the language translation software. The commercial feasibility of this design will also be investigated. This includes identifying potential applications, languages to be supported, cost, and user requirements such as interface modes and response times. By combining both the commercial and technical elements, a complete definition of successful software and system solutions for pocketable language translation devices will be achieved.

Prototype systems showing device independence will be developed and demonstrated and a final report written documenting the Phase I results and recommendations for follow-on research and development in Phase II. Options are included for incorporating additional language pairs into the system and application specific terminology.

Anticipated Benefits/Potential Commercial Applications of the Research or Development:

Applications include all individuals who require multi-lingual capabilities. The mobile translator will benefit a wide range of individuals including military personnel, airport employees, border patrol and customs agents, police, fire fighters, retail clerks, bank tellers, delivery personnel, phone operators, tourists and any industry that sells, develops or manufactures products to/in global markets or employs individuals that do not speak the native language.

A.4 Camera-Based Mode

A4.1 -- Brief Summary

The primary means to input text into the SmartPhone for this mode of usage will be a digital camera. Such a system is shown in Figure CM1. The digital camera will be used to capture an image of the foreign language. Such a picture is shown in Figure CM2.

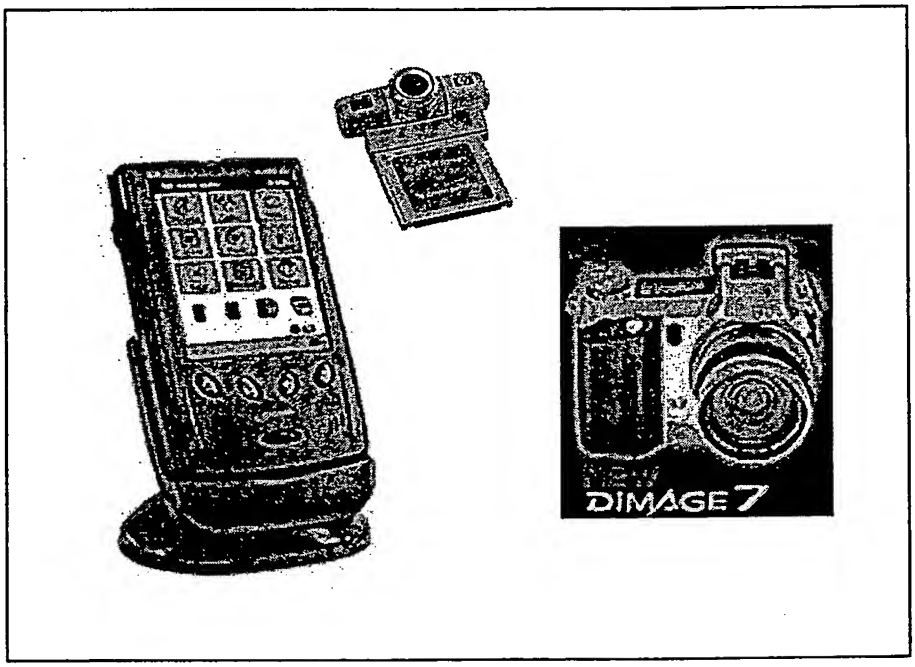

Figure CM1: Examples of Camera-based Systems

Once the desired image is obtained, the SmartPhone will wirelessly connect to a remote server where the image will be processed and the resulting translation sent back to the user. An example of the translated text in the proposed "one-click" GUI is shown in Figure CM3. For most applications, this connection will be made using cellular telephones. Because of the limited bandwidth of such a connection, it is important to reduce the overall size of the transmission. Thus, SpeechGear is in the process of evaluating different image compression algorithms. These algorithms will be embedded directly into SpeechGear's software, and thus will be transparent to the end user.

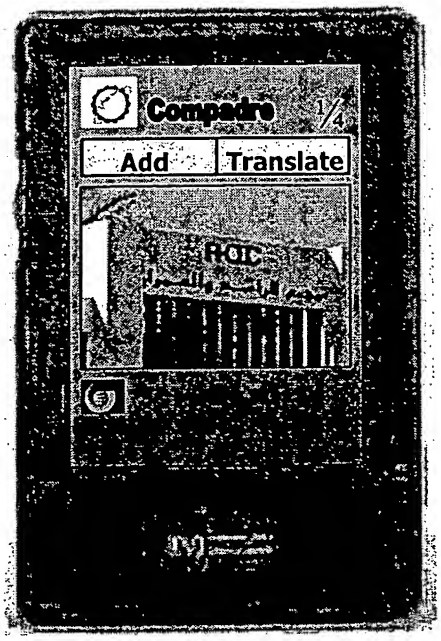

Figure CM2: Example Graphical User Interface to Submit Images for Translation

As is shown in Figure CM2, a "one-click" GUI is planned. After capturing the image(s), the user will simply select "Translate" and the wireless connection will automatically be established. Note that multiple images can be sent simultaneously using a single click. This is similar to the "Add to Basket" interfaces that are being used at web-based shopping sites. In this approach, items that are selected can be loaded into a virtual basket or cart, and once you are done shopping you can select "Check Out" to purchase all of the items simultaneously. For *Compadre*, multiple images can be selected and entered into the queue, and when the user is ready to connect to the remote server, then simply selecting the "Translate" button will connect the SmartPhone to the remote server, which in turn will process the images and return the resulting translation. The images will be transmitted back to the user using an HTML format. The users can then scroll through these images and save or delete them as is desired. Please note that the actual buttons will be Icons versus text, and thus the look and feel of the resulting GUI will be a substantial improvement over what is shown in the Figures.

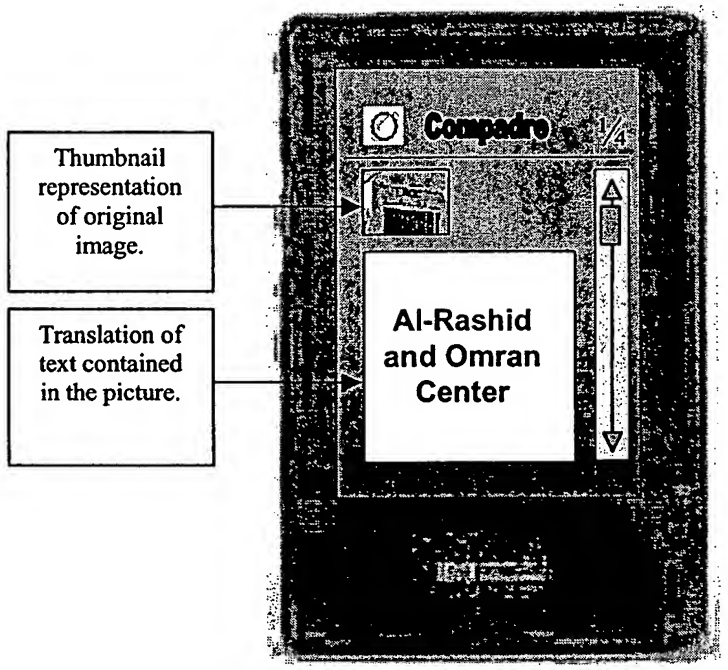

Figure CM3: Example of Graphical User Interface for Viewing Results of Translation

A4.2 – System Requirements:

CM1. System shall support connectivity to a digital camera (1.0a).

CM2. System shall be capable of displaying the captured digital image (1.0a).

CM3. System shall allow the user to select region of text to be translated (1.0c).

CM4. System shall allow the user to use "One-Click" to transfer image to remote server for processing (1.0b).

CM5. System shall allow the user to add multiple images to the send buffer (1.0c).

CM6. System server shall be capable of processing multiple images on a single connection/transmission (1.0c).

CM7. System shall support wireless connectivity such as a cellular telephone (1.0a).

CM8. System shall include image compression algorithms to reduce transmission connect time (1.0c).

CM9. User shall have the ability to turn on/off the image compression capability (1.0c).

CM10. The returned image shall include a "thumbnail" picture of the original image along with the translated text (1.0a).

CM11. The user shall be capable of saving this return image on the SmartPhone (1.0b).

CM12. The user shall be capable of scrolling through multiple return images using a "one-click" interface (1.0b).

-- Note: See Table CM1 for a summary of requirements CM13 through CM44.

CM13. The system shall support bi-directional translation for English/Arabic (1.0a).

CM14. The system shall support bi-directional translation for English/Korean (1.0a).

CM15. The system shall support bi-directional translation for English/Japanese (2.0a)

CM16. The system shall support bi-directional translation for English/Spanish (2.0b)

CM17. N/A

CM18. The system shall support bi-directional translation for English/Serbian (2.0c)

CM19. The system shall support bi-directional translation for English/Mandarin Chinese (2.0c)

CM20. The system shall support single-directional translation for Mandarin Chinese to English (2.0b)

CM21. The system shall support single-directional translation for Serbian to English (2.0b)

CM22. The system shall support bi-directional translation for English/Albanian (3.0b)

CM23. The system shall support single-directional translation for Albanian to English (2.0b)

CM24. The system shall support bi-directional translation for English/Thai (2.0c)

CM25. The system shall support single-directional translation for Thai to English (2.0b)

CM26. The system shall support bi-directional translation for English/Creole (3.0b)

CM27. The system shall support singledirectional translation for Creole to English (2.0b)

CM28. The system shall support bi-directional translation for English/ Indonesian (2.0c)

CM29. The system shall support single-directional translation for Indonesian to English (2.0b)

CM30. The system shall support bi-directional translation for English/ Turkish (2.0c)

CM31. The system shall support single-directional translation for Turkish to English (2.0b)

CM32. The system shall support bi-directional translation for English/Malay (2.0c)

CM33. The system shall support single-directional translation for Malay to English (2.0b)

CM34. The system shall support bi-directional translation for English/Greek (2.0c)

CM35. The system shall support single-directional translation for Greek to English (2.0b)

CM36. The system shall support bi-directional translation for English/Russian (2.0c)

CM37. The system shall support single-directional translation for Russian to English (2.0b)

CM38. The system shall support bi-directional translation for English/French (2.0a)

CM39. The system shall support bi-directional translation for English/German (2.0a)

CM40. The system shall support bi-directional translation for English/Portuguese (2.0b)

CM41. The system shall support bi-directional translation for English/Hindustani (2.0c)

CM42. The system shall support single-directional translation for Hindustani to English (2.0b)

CM43. The system shall support bi-directional translation for English/Swedish (2.0b)

CM44. The system shall support bi-directional translation for English/Norwegian (2.0c)

Table CM1: Summary of Language Support Schedule

Language	Single Directional	Bi-Directional
Arabic	CM13 – 1.0a	CM13 – 1.0a
Korean	CM14 – 1.0a	CM14 – 1.0a
Japanese	CM15 – 2.0a	CM15 – 2.0a
Spanish	CM17 – 2.0b	CM16 – 2.0b
Serbian	CM18 – 2.0c	CM18 – 2.0c
Mandarin Chinese	CM20 – 2.0b	CM19 – 2.0c
Albanian	CM23 – 2.0b	CM22 – 3.0b
Thai	CM25 – 2.0b	CM24 – 2.0c
Creole	CM27 – 2.0b	CM26 – 3.0b
Indonesian	CM29 – 2.0b	CM28 – 3.0b
Turkish	CM31 – 2.0b	CM30 – 3.0b
Malay	CM33 – 2.0b	CM32 – 3.0b
Greek	CM35 – 2.0b	CM34 – 3.0b
Russian	CM37 – 2.0b	CM36 – 2.0c
French	CM38 – 2.0a	CM38 – 2.0a
German	CM39 – 2.0a	CM39 – 2.0a
Portuguese	CM40 – 2.0b	CM40 – 2.0b
Hindustani	CM42 – 2.0b	CM41 – 2.0c
Swedish	CM43 – 2.0b	CM43 – 2.0b
Norwegian	CM44 – 2.0c	CM44 – 2.0c

Note: If a discrepancy is present between the table entries and the line items, the line items take precedence.